A Program for Estimating Age Specific Mortality

Daniel Kaplan

Sept. 24, 2010

Given a set of birth and death years for individuals, find the age specific mortality at different ages. For an example, and for debugging purposes, consider the following birth and death years:

```
> npeople = 1e+05
> lastyear = 2010
> birthyear = runif(npeople, min = 1900, max = lastyear)
> lifespan = (rexp(npeople, rate = 0.05) + runif(npeople, min = 0,
+     max = 20) + runif(npeople, min = 0, max = 20))
> infantdeaths = which(runif(npeople) < 0.3)
> lifespan[infantdeaths] = runif(length(infantdeaths), min = 0.05,
+     max = 0.25)
> deathyear = birthyear + round(lifespan)
> deathyear[deathyear > lastyear] = NA
```

The basic structure will be a set of starting years, say 0, 1, 5, 10, 15, 20, ∞, and a count for each starting year: whether the person reached that start year and whether they died before the next starting year. For each person, look to see if they reached the start year. If so, add one to the **reached** variable. If they died before the next start year, add one to the **died** variable.

In addition, for each start year, there is a lockout. The person had to be born after the lockout year in order to count them toward that starting year. The point of this is to avoid using very old people to infer mortality for very young people. This is because conditions might have changed substantially.

```
> startyears = c(0, 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50)
> lockouts = -Inf * c(2000, 2000, 1990, 1985, 1975, 1970, 1965,
+     1960, 1955, 1950, 1940, 1920)

> mortality.count = function(by, dy, startyears = 0:100, lockouts = rep(-Inf,
+     length(startyears)), currentyear = 2010) {
+     age = dy - by
+     endyears = c(startyears[2:length(startyears)], Inf)
+     reached = rep(0, length(startyears))
+     died = rep(0, length(startyears))
+     for (k in 1:length(startyears)) {
```

```
+            aliveinds = which(((age >= startyears[k]) | (is.na(dy) &
+                ((currentyear - by) >= startyears[k]))) & by > lockouts[k])
+            reached[k] = length(aliveinds)
+            if (length(aliveinds) > 0) {
+                if (endyears[k] == Inf)
+                    died[k] = length(aliveinds)
+                else died[k] = sum(age[aliveinds] < endyears[k],
+                    na.rm = TRUE)
+            }
+        }
+        return(data.frame(sy = startyears, ey = endyears, n = reached,
+            d = died))
+ }
```

To run the program:

```
> foo = mortality.count(birthyear, deathyear, startyears, lockouts)
```

By default, the startyears are every year, and the lockouts are disabled.

```
> foo = mortality.count(birthyear, deathyear)
> plot(foo$sy, 1000 * foo$d/foo$n, xlab = "Age (years)", ylab = "Mortality rate (per 1000)")
```